



# РОМИП 2006: отчет организаторов

---

*И. Некрестьянов  
М. Некрестьянова*

---

# Что такое РОМИП?

---

## **Российский семинар по оценке методов информационного поиска**

- Русскоязычные задания
- Крупные коллекции
- Использование апробированных подходов к оценке
- Равноправие и анонимность участников
- Использование независимых экспертов для оценки результатов поиска
- Возможность повторного использования результатов

# Что требуется от участника?

---

- Подать заявку
- Участвовать в формировании правил проведения дорожек
- Выполнить полученные задания и сдать ответы в оргкомитет
- Проанализировать результаты оценки и подготовить статью
- Компенсировать часть затрат на проведение оценки и организацию семинара
- Сделать доклад на очном семинаре

# Как производится оценка?

---

- Задания для оценки отбираются оргкомитетом ПОСЛЕ сбора ответов участников
- Оценка на основе сравнения с «эталоном»
- «Эталон» обычно строится с помощью ассессоров, которые оценивают ответы на правильность
- При повторении дорожки часть заданий оценивается повторно
- По результатам вычисляются некоторые стандартные метрики

# Построение «эталона»

---

- Ассессор НЕ знает какой системы это ответ
- Один ассессор оценивает ответы разных систем
- Порядок выдачи заданий для оценки случаен
- Обычно минимум два ассессора оценивает одно и тоже задание
- Многозначная шкала (4 значения)
- Оцениваются *уникальные* ответы (экономия)
- Использование расширенных описаний (Цель – упразднить неоднозначность трактовки, понять точку зрения ассессора)

# Коллекции РОМИП

---

	Состав	Размер	Предоставлена
Narod.Ru	Веб-сайты из домена narod.ru	728.000 док. 22 000 сайтов.	Яндекс
Legal	Законодательство РФ	60.000 док.	Кодекс
DMOZ	Веб-сайты из русской части DMOZ	300.000 док. 2087 сайтов	Рамблер
<b>News</b>	Все новости за три периода из 17 источников	<b>31.500 док.</b> <b>75 Мб.</b>	Яндекс

# Дорожки по поиску

---

	Заданий	Оценивалось	Объем оценки	Согласие асессоров
Поиск по Веб	24595	70 (20)	340	<b>0.94</b>
Поиск по норм. коллекции	12925	50 (0)	130*	-
Поиск по смешанной коллекции	37091	158	-	<b>0.94</b>
Поиск по образцу	11358	<b>233</b>	-	-

# Дорожки по поиску (2)

---

- Совместная оценка для разных дорожек
- Задания для оценки отбирают ассессоры
  - Запросы из  $\geq 2$  слов
  - «Расширенное описание» - представление ассессора о цели поиска.
- Экспертная оценка для нормативных запросов
  - Только 1 оценка на задание. Выборочная проверка.
- «Поиск по образцу»
  - Аналогично РОМИП'2005
  - Запросы из оценивавшихся в 2004-2005 году
  - Образцы – релевантные документы (по схеме AND)
  - Оценка: проверка ответов на соответствие *запросу*



# Тематическая классификация

	Категорий	Обучающее множество	Оценивалось	Объем оценки	Согласие ассессоров
Веб сайтов	247	2116 сайтов	24 кат. 2906	110	0.90
Веб докум.	247	2116 сайтов	24 кат. 7695	200	0.88
Нормат. докум.	183	13772 док.	40 кат.	0	-

- Выборочная оценка Веб страниц (10%, <75 док.)
- Доп. оценка для дорожки классификации Веб страниц
- **Закончилось эталонное множество для нормат. док.**

# Аннотирование

---

**Задание:** запрос + документ

**Ответ:**

фрагмент текста < 300 символов без HTML разметки

**Проверяем:**

по аннотации можно предсказать релевантность документа.

**Ассессор:**

- видит аннотации по одной
- оценивает равное число ответов разных систем
- НЕ видит аннотации одного и того же от разных систем.

**Эталон:**

результаты оценки дорожек поиска 2004-2005

# Аннотирование (оценка)

---

- Повторение по правилам прошлого года
  - Также смесь заданий:  
50% сильно-, 25% слабо и 25% нерелевантных
- Всего 2 прогона от участников
- **Разница в итоговых оценках достигает 10%**  
(в прошлом году оценки были очень схожи)
- Общая проблема:
  - Результаты оценки невозможно повторно использовать.

# Вопросно-ответная дорожка

---

**Основа:** Narod.Ru

**Задания:** по 200 заданий от участников, 215 оргкомитет

*Кто родился в год змеи?*

*Почему украинцы воруют российский газ?*

*На чем ездят россияне?*

**Ответ системы (до 10):**

фрагмент текста до 300 символов без html разметки

**Оценка:**

- Все 615 заданий, 2 ассессора
- Ассессор видел сразу **все** варианты ответов
  - Случайный порядок, без указания системы
- Для 30-40% нет релевантных ответов

# Шкала оценки

---

- Нет ответа
- Точный ответ
  - государства Америки, открытой в 1492 году Христофором Колумбом
  - Америку за семьдесят лет до Колумба открыл китайский мореплаватель Чжэнь Хи
- Частичный ответ
  - Стоит напомнить, что в год Змеи родились Коперник, Николай Кузанский, Циолковский.
  - На машинах этой марки ездят десятки тысяч россиян, потому весть о трудностях фирмы вызвала понятную нервозность у обладателей "НЕКСИЙ" и "Тико".
- Где-то рядом
  - А В год Обезьяны родились : А Македонский... А А ГОД ЗМЕИ...
  - Этот автомобиль позволил воплотить в реальность мечту многих россиян ездить на новой иномарке

# Организация новостей

---

Задача: структурировать поток новостей в набор событийных сюжетов, связанных между собой.

Событие: отражение события из реального мира в потоке новостных сообщений  
(«смерть папы римского»)

Сюжеты – события, связанные между собой нетематическими связями  
(«перенос футбольного тура в связи с кончиной папы»)

- 2006: 2 участника, 4 прогона

# Подходы к оценке

---

- Позиция “редактора”
  - Полная разметка новостного потока вручную
  - Хорошо позволяет оценить полноту (и точность)
  - Легко повторно применяемо
  - Но проверяется ответ для “куда бы я бы положил это а месте редактора”
- Позиция “читателя”
  - Проверяем насколько хороши результаты системы
    - Нет ли мусора в выделенном событии/сюжете?
  - Трудно оценить полноту и переиспользовать
  - Но ближе реальному пользователю
- Гибридные подходы

# Инструмент асессора

The screenshot shows the Mozilla Firefox browser window with the following elements:

- Browser Title Bar:** Mozilla Firefox
- Menu Bar:** Файл, Правка, Вид, Переход, Закладки, Инструменты, Справка
- Left Panel (Tree View):**
  - добавить сюжет [go to tasks](#)
  - Долго зрела ягодка "Калина". Перезр
  - Долго зрела ягодка "Калина". Перезр
  - Долго зрела ягодка "Калина". Перезр
  - АвтоВАЗ начинает производство "Кали
  - ВАЗ начинает промышленное производ
  - Сегодня с конвейера АвтоВАЗа выйдет
  - добавить событие
  - Над Суданом взорвался Ан-12 с грузо
  - Над Суданом взорвался Ан-12 с грузо
  - Над Суданом взорвался Ан-12 с грузо
  - В Судане взорвался самолет с деньга
  - Над Суданом взорвался самолет с ден
  - Авиакатастрофа в Судане: разбился с
  - Во взорвавшемся Ан-12 граждан Росси
  - Во взорвавшемся Ан-12 граждан Росси
  - Среди погибших членов экипажа и пас
  - Авиакатастрофа в Судане: последние
  - Авиакатастрофа в Судане: последние
  - добавить событие
  - 21 ноября "Российская газета" опубл
  - Тюменская ТЭЦ-1 готовится к пуску
- Main Content Area:**
  - 18.11.2003 9:41:04 Предприниматели Миасса голосуют за принцип "одного окна"
  - 18.11.2003 9:41:04 Челябинская область погасила долг по детским пособиям на 96%
  - 18.11.2003 9:41:04 В Озёрске ведётся работа по реорганизации муниципальных предприятий
  - 18.11.2003 9:41:18 В Москве расстреляли директора угольной компании
  - 18.11.2003 9:41:19 Окружные ревизоры проверили нефтеюганские школы
  - 18.11.2003 9:50:15 Впервые за 12 лет на Таймыре пройдет съезд местных оленеводов
  - 18.11.2003 9:50:15 В Петербурге 18 и 19 ноября будут обсуждать инвестиционные проекты
  - 18.11.2003 9:50:15 Выставка "Гибкий стебель" открылась в модном доме "Ulita"
  - 18.11.2003 9:50:16 Виагра для Путина
  - 18.11.2003 9:50:16 В Благовещенске традиция...
  - 18.11.2003 12:40:49 У коммунистов плохая память на дорожные номера**
  - В понедельник Центризбирком обнаружил информацию о недостоверных данных и сведения, которые соискатели депутатских мандатов по каким-то причинам не указали в своих декларациях. Мало того, что целых 60 товарищей из КПРФ плохо знают свое имущественное состояние, но еще поражают объемы этой забывчивости. Сразу шесть кандидатов-коммунистов "забыли" о принадлежащих им "Мерседесах". Среди них депутаты Госдумы Валентин Никитин и Геннадий Гамза, а кандидат Василий Алтухов, упомянув о принадлежащих ему "Волге" и тракторе Т-25, так и не вспомнил про "Мерседес-Бенц S600L" (1998 г.в.). В списке "забывчивых" от КПРФ пестрят автомобили "БМВ", "Ниссан", "ВАЗ" и "УАЗ". Да и с доходами не все в порядке. Бывший генпрокурор Юрий
- Right Panel (List View):**
  - 20:25 Около 20 домов в Пер...
  - 12:721 ноября "Российска
  - 21:44 Что делает Америка
  - 23:34 Началось...
  - 0:3 Невазлин больше не ре
  - 18:537 рублей за проезд в
  - 22:35 Лондон встречает Буш
  - 18:53 Администрация г. Вла
  - 22:45 "Вашингтонский снайп
  - 0:34 Россия хочет передат
  - 19:13 Владимир Потанин ста
  - 23:44 Евро (ЦБ, 19 ноября)
  - 22:35 К приезду Буша в Лон
  - 23:54 Рост ВВП в РФ в янва
  - 23:33 Рост ВВП за 10 меслц



# Заметки об оценке

---

- Что оценивалось
  - все сообщения за первые два дня для каждой недели
  - ~3500 в каждой неделе
- Планировавшаяся оценка завершена, но по-видимому требуется дополнительная работа
- Первые (неожиданные) наблюдения:
  - мало выделенных кластеров в первых результатах
  - распределение кластеров не совпадает с ожиданиями
  - сильные отличия у разных ассессоров
  - внутренняя структура сюжетов часто недоработана (очевидные дубли не склеены)

# Возможные причины

---

- Модификации инструмента в процессе оценки
  - Исправление найденных ошибок
- (Возможно) не очень четкое понимание задачи ассессорами
- Техническая необходимость расщепления ленты на фрагменты по 500-1000 сообщений
- **Очень сложная задача для ассессора**
  - Выбор между сотнями готовых кластеров
  - Невозможно быстро понять о чем этот сюжет не просмотрев сообщения еще раз
  - Контекст быстро забывается, если ассессор отвлекся на что-то другое

# Что мы планируем

---

- Работа над ошибками
  - Попросить ассессоров перепроверить все нетривиальные сюжеты
- Изменить цель оценки - вместо “разложи все по кластерам” перейти к
  - Найти максимум нетривиальных сюжетов
  - Чем крупнее сюжет, тем лучше
- Технические изменения
  - Одиночные сюжеты остаются в списке
  - Сообщения можно/нужно просматривать не последовательно (использовать поиск)

# Особенности РОМИП'2006

---

- 👍 Появилось НП РОМИП
  - Возможность принимать оргвзносы по безналу
  - **Есть кому передавать права на коллекции**
- 👉 Закончилось идеальное множество для классификации нормативных документов
- 👍 Новая интересная вопросно-ответная дорожка
- 👍 Реальный запуск дорожки по структуризации новостей
- 👉 Большое число «самоотводов» в последний момент